170-WP-007-005

CEOS Interoperable Catalogue System System Design Document

ECS White Paper

June 1998

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Janet Hylton /s/ 6/4/98

Janet Hylton Date
EOSDIS Core System Project

SUBMITTED BY

George Percivall /s/ 6/4/98

George Percivall Date
EOSDIS Core System Project

Raytheon Corporation Systems
Upper Marlboro, Maryland

CEOS

Committee on Earth Observation Satellites
Working Group on Information Systems and Services

Interoperable Catalogue System (ICS)
System Design Document (SDD)

CEOS
Working Group on Information Systems and Services
Protocol Task Team

Doc. Ref.: CEOS/WGISS/PTT/SDD
Date: June 1998
Issue: Version 1.4

AUTHORITY

Issue:	Version 1.4
Date:	June 1998

This document has been approved for publication by the Protocol Task Team of the Committee on Earth Observation Satellites (CEOS) and reflects the consensus of the Protocol Task Team technical panel experts from the CEOS member agencies.

This document is published and maintained by:
Goddard Space Flight Center
National Aeronautics and Space Administration
http://edhs1.gsfc.nasa.gov

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DOCUMENT STATUS SHEET

Version	Date	Comments
0.1	May 1996	Interoperable Catalogue Systems and Services System Architecture Document (ECS White Paper 170-WP-007-001). Annotated Outline of SAD prepared for PTT Meeting, May 1996, Tokyo.
1.0	September 1996	Document renamed ICS SDD. RIDs from Tokyo PTT meeting incorporated. Document prepared for PTT Meeting, September 1996, Annapolis. Final scope and purpose. Initial technical completeness review.
1.1	November 1996	RIDs from Annapolis PTT meeting incorporated. Document prepared for PTT Meeting, December 1996, Ispra. Document to be reviewed by PTT using formal RID process.
1.2	March 1997	RIDs from Ispra PTT-4 meeting incorporated. Document prepared for ratification by PTT.
1.3	February 1998	Document prepared for RID review during PTT.
1.4	June 1998	RIDs from Salzburg PTT-8 meeting incorporated. Document prepared for ratification by PTT.

DOCUMENT CHANGE RECORD

Change No.	Date	Changes in Version 1.4 from Version 1.3
1	March 1998	Incorporated RIDs from PTT-8 meeting in Salzburg.

1. INTRODUCTION

1.1 SDD Purpose and Scope

ICS Compatibility: Explanatory

Major Space Agencies through the Committee for Earth Observation Satellites (CEOS) have recognized the importance of providing a seamless access to their catalogues via a new Interoperable protocol. This protocol is the Catalogue Interoperability Protocol (CIP) and is defined in a sister document: the CIP Specification - Release B. The CIP Specification defines how compliant catalogues need to interface through the low level Z39.50 protocol. This System Design Document (SDD) focuses on the design of a global system and the infrastructure that will link together agencies using the CIP protocol. This macro system is called the Interoperable Catalogue System or ICS.

The CEOS Protocol Task Team (PTT) is developing a suite of documents as defined in the PTT Development Plan [R1]*. The ICS User Requirements Document (URD) [R2] is where the CEOS agencies bring in their needs for functionality, constraints, etc. The PTT accepts input as requirements added to the URD. It is the PTT activity to interpret those requirements into the CIP Specification [R3], the ICS Guide Design and Protocol Specification [24], and the ICS SDD. As a result, if groups comply with the CIP Specification, the ICS Guide Design and Protocol Specification, and the SDD, then they will be compliant with the URD. This section defines the purpose of the SDD. The purpose of ICS is provided in Section 2.

The ICS SDD defines the elements and interfaces which comprise the CEOS ICS. Implmenentors of ICS elements should derive element requirements from the mandatory parts of the ICS System Design. This document defines relationships to the CEOS Network which is under development by the CEOS Network Subgroup. Assumptions are made in this document about existing systems which interface to ICS, e.g., existing agency systems will persistently store user orders. These assumptions cannot be required of the agency systems but represent the SDD's approach to the interface with ICS. The last section of the SDD defines a mimimum site configuration relative to the SDD content in order to meet CEOS policy. The SDD is written for a federation of CEOS agency systems, however other data provider federations could use the SDD as a template design for their system. The ultimate purpose of the SDD is to define a system which can be implemented and operated by the CEOS federation to provide data and services to users.

* Citations for references are provided in Section 1.5.

Each of the lowest level sections in the SDD are labeled for ICS compatibility. The applicability labels tell the reader if an ICS site must implement the content of the section in order to be compatible with other ICS sites. The determination of applicability of a section is based on adherence to CEOS policy, effective ICS operation, and assurance of interoperability. The following terms are used in the applicability labels.

- Mandatory - A site must comply with contents of the section to be ICS compatible.
- Mandatory As Applicable (MAA) - If a site is implementing the functionality in the section, then the site must comply with the section to be ICS compatible. A minimum site need not comply with any MAA section and could still be ICS compatible.
- Explanatory - The contents of the section describe a concept or background topic which may be useful to the reader to understand ICS. ICS compatibility is not determined by the contents of an explanatory section.

1.2 Organization of the SDD

ICS Compatibility: Explanatory

The SDD is organized into various views of the ICS. This approach is based on guidelines for SDDs, e.g., [R6] and existing SDDs, e.g., [R7]. After a discussion of the PTT approach to catalogue interoperability (Section 2), the SDD provides the following views of the ICS:

- Functional Framework (Section 3) - overall application view of the computing elements which comprise the functional architecture of ICS. The ICS Framework provides several configurations for ICS implementation at sites.
- Data View (Section 4) - provides an overall data framework indicating what data is held by the various ICS elements. Several data components are defined in the Retrieval Manager in order that ICS operations are performed correctly across ICS sites.
- Communications View (Section 5) - describes two models of the ICS design related to communications and networks. First a communication protocol stack is defined placing CIP on a TCP/IP stack. Next, ICS reliance on CEOSnet is discussed including identifying sites at which ICS Retrieval Managers are initially intended to be located and the traffic between the sites.
- Security View (Section 6) - This section provides a discussion of the security design for ICS. First, a security assessment of ICS is provided in terms of security assets, vulnerabilities and threats. Based on the assessment, the ICS security control methods are described. The methods are grouped into three categories: administration, physical and computing controls.
- Systems Management View (Section 7) - This section provides a list of topics which can serve as a basis for beginning the System Management design in the next ICS release.
- Architecture Verification (Section 8) - This section provides several methods for evaluating the system design. Through a series of scenarios, a definitive identification of inter-element interfaces, and distributed query performance estimates, this section demonstrates the completeness of the system design presented in earlier sections.
- ICS Minimum Site Configuration (Section 9) - This section defines the required minimum configuration for a site which wishes to be considered ICS compatible. The minimum site is based on the CEOS policy.

1.3 ICS/PTT Development Process

ICS Compatibility: Explanatory

This document was developed by the CEOS PTT. The PTT is part of the WGISS Access Subgroup within CEOS. The lead agency for compilation of PTT inputs and for final preparation of this version of the document was the National Aeronautics and Space Administration (NASA). A complete list of organizations participating in the PTT is provided in the PTT Terms of Reference (http://ceos.ccrs.nrcan.gc.ca/taskteam/cip.html)

1.4 A Guide To PTT Documents

ICS Compatibility: Explanatory

The CEOS Protocol Task Team (PTT) collects requirements for an interoperable infrastructure connecting Earth-observation data catalogues and access systems. It establishes agreement on these requirements and produces detailed specifications on which system implementations can be based. The PTT has named this interoperable infrastructure the CEOS Interoperable Catalogue System (ICS). The PTT output is captured and maintained in the documents listed below. A discussion of which document a user might read first is provided after the list of PTT documents.

- **ICS User Requirements Document (URD)**

  The ICS URD [R2] specifies user requirements for an interoperable infrastructure linking catalogue systems of different agencies. For this purpose it defines requirements for all interoperable components and the protocols needed for exchanging messages between them. (Currently, the URD does not reflect the http approach to guide and will be updated.)

- **ICS System Design Document (SDD)**

  The ICS SDD defines the elements and interfaces which comprise the CEOS Interoperable Catalogue System (ICS). The SDD provides diagrams showing the interrelations between ICS elements, scenarios to explain the dynamic interaction, a data model showing the data relations, the communications services utilized in ICS, and the system management approach for ICS. The SDD provides both design and tutorial information.

- **Catalogue Interoperability Protocol (CIP) Specification - Release B**

  The CIP Specification [R3] defines the interoperable protocol for exchanging messages related to data search and data ordering. CIP is defined as a profile of the ISO standard Z39.50 with extensions for distributed searching using the collections model. The specification defines all CIP messages, as well as the attributes used for searching and the elements needed for retrieval. CIP may be used outside of the CEOS ICS. The CIP Specification is the definitive

source for determining CIP compliance. The CIP Specification provides the framework for exchanging data orders. Details on how to specify options on orders are defined in the Order Options Amendment to the CIP Specification [R28].

- **ICS Guide Design and Protocol Specification**

  The ICS Guide Design and Protocol Specification [R24] describes the ICS elements that support guide and the protocol of messages used for guide. The approach is based on the http protocol using virtual documents. The interaction of the Guide elements with CIP elements is discussed. Example scenarios describe the dynamic behavior. The ICS Guide Design and Protocol Specification is the definitive source for determining ICS Guide Protocol (IGP) compliance.

- **ICS Collection Manual**

  The ICS Collection Manual [R5] provides procedures and guidelines for the creation and maintenance of Collection Information contained in an ICS Retrieval Manager. The document provides sufficient detail to allow the ICS Site Administrators to manage the ICS Collection repositories according to the rules specified in the Collection Manual. The manual further provides guidelines for developing the ICS Collection Structure. Collectively, the procedures and guidelines; which can be applied to any/all implementation strategies, if followed, will ensure data interoperability.

- **ICS Valids Document**

  The ICS Valids Document [R4] defines the list of valid keywords for the enumerated search attributes used by CIP. The valids document provides the procedures for controlling the list of valids either based on coordination with other standardization groups or through rules and procedures for ICS only valids

A starting point for most users of PTT documents will be the SDD which provides tutorial information about how CIP, IGP and Collections are used in ICS. An implementer who wants the details of CIP messages may want to go directly to the CIP Specification. This is also true for an IGP implementer who may want to go directly to the Guide Design and Protocol Specification. Someone who is responsible for organizing the data for an agency may want to browse the SDD to understand the ICS data model and then proceed to the details in the Collections Manual and the Valids Document. If technical input to the PTT direction is desired, reviewing the URD and proposing new User Requirements is the right approach.

To avoid conflicting information, the overlap of material between PTT documents has been minimized. Specific guidelines for splitting information between the SDD and the CIP Specification were used and are presented in Table 1-1.

Table 1-1. Split of Information between CIP Specification and ICS SDD

Dimension	CIP Specification	ICS SDD
Scope	Application Protocol	Full Protocol Stack System Management
Data	Internal content of CIP messages	CIP Messages by Name only
Element Interactions	Session between a Target and Origin	Use of multiple CIP sessions across multiple ICS Elements

Additional information about PTT activities and documents can be found at:
http://ceos.ccrs.nrcan.gc.ca/taskteam/cip.html

1.5 Glossary

1.5.1 Acronyms

ICS Compatibility: Explanatory

The following acronyms are used in this document:

ADD	Architecture Design Document
ANSI	American National Standards Institute
APDU	Application Protocol Data Unit
AS	Access Subgroup (Part of CEOS-WGISS)
ASN.1	Abstract Syntax Notation.1
BER	Basic Encoding Rules
BNSC	British National Space Centre
CA	Certification Authority
CCRS	Canada Centre for Remote Sensing
CCSDS	Consultative Committee for Space Data Systems
CDB	Collection Data Base
CEO	Centre for Earth Observation (European Commission)
CEO-ES	CEO- Enabling Services
CEOS	Committee on Earth Observation Satellites
CINTEX	Catalogue INTeroperability Experiment
CINTOPS	Catalogue INTeroperability OPerationS
CIP	Catalogue Interoperability Protocol
CM	Configuration Management
CMT	Collection Management Tool
CNES	Centre National d'Etudes Spatiales (France)

CSIRO	Commonwealth Scientific and Industrial Research Organisation (Australia)
DB	Data Base
DBMS	Data Base Management System
DEM	Digital Elevation Map
DIF	Directory Interchange Format
DLR	Deutsche Forschungsanstalt fur Luft-und Raumfahrt
DNS	Domain Name Service
DOGS	Directory of GEO Servers
ECS	EOSDIS Core System
EO	Earth Observation
EOC	Earth Observation Center (NASDA)
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System (NASA)
ESA	European Space Agency
ESRIN	European Space Research Institute (ESA)
FGDC	Federal Geographic Data Committee (USA)
GCMD	Global Change Master Directory
GI	Guide Indexer
GS	Guide Serv er
GSFC	Goddard Space Flight Center (NASA)
GT	Guide Translator
HITS	Hughes Information Technology Systems, Inc.
HTML	Hyper Text Mark-up Language
HTTP	HyperText Transfer Protocol
ICS	Interoperable Catalogue System
IGP	ICS Guide Protocol
ID	Identifier
IP	Internet Protocol
IRE-RAS	Institute of Radio Engineering. - Russian Academy of Science
ISA	ICS Site Administrator
ISO	International Standards Organization
LaRC	Langley Research Center (NASA)
LDAP	Lightweight Directory Access Protocol
MAC	Message Authentication Code
MACAO	Member Agency Control Authority Office
MCT	Monitoring and Control Tool
NASA	National Aeronautics and Space Administration (US)
NASDA	National Space Development Agency (Japan)
NOAA	National Oceanic and Atmospheric Administration (US)
NRSC	National Remote Sensing Center
NSRS	Natural Environment Research Council (BNSC)
OHS	Order Handling System
OSI	Open Systems Interconnection

PTT	Protocol Task Team (Part of CEOS- WGISS-AS)
PKI	Public Key Infrastructure
QA	Quality Assurance
RFC	Request For Comment
RM	Retrieval Manager
RPN	Reverse Polish Notation
SATAN	Security Administrator Tool for Analyzing Networks
SDD	System Design Document
SFDU	Standard Formatted Data Unit
SSM	Site System Management
SNMP	Simple Network Management Protocol
SST	Sea Surface Temperature
TBD	To Be Determined
TBR	To Be Resolved
TBS	To Be Supplied
TCP	Transmission Control Protocol
TN	Technical Note
UDB	User Data Base
UPS	User Profile System
UR	User Requirement
URD	User Requirements Document
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
US	United States
USGS	U.S. Geological Survey
WGISS	Working Group on Information Systems and Services (Part of CEOS)
WWW	World Wide Web

1.5.2 Definitions

ICS Compatability: Explanatory

This section provides definitions of the terms related to ICS:

Archive	An archive of EO data can hold various types of data ranging from satellite images and climatological products processed from the images, to observation data and climatological statistics. An archive may also contain information describing the EO data and also supplementary data such as design documentation, algorithm object and source code, technical reports, user manuals, etc. There is likely to be a database management system for maintenance and low level access to the data. The archive will, in general, be accessed by a front end archive server that then presents the data as requested by the Retrieval Manager.
Catalogue Interoperability	The ability to provide a Data User with the appearance of a single, unified catalogue for all participating data providers. In order to provide catalogue interoperability all paricipating data providers must support at least one common method (i.e., API) for accessing functions such as authentication, directory, inventory, guide and order. Each supplier may support additional consumer functional interfaces to support their private data users
Catalogue System	A catalogue system provides services such as inventory, browse, directory, order and guide, which may be supplemented by further services, but should contain at a minimum, inventory. The CIP is the protocol that shall enable the many services of many catalogue systems to inter-operate. Usually a catalogue system resides at a particular agency or data provider facility but may be distributed across catalogue sites.
Catalogue Translator	One of three types of ICS *Translators*. *Catalogue Translator* converts CIP messages into a data providers protocol for the services of Inventory, Directory, and Browse.
CEOS Agency Systems	The data provider systems of CEOS member agencies and their affiliates.
CIP Client	A software element composed of a *Presentation Layer*, a *Local User Management Layer*, and an *Application Layer*. Only the *CIP Client Application Layer* is part of ICS.
CIP Client Application Layer	Part of the *CIP Client* which deals directly with CIP including creating CIP messages and includes off-the-shelf Z39.50 communication software.
CIP Client Local User Management Layer	Part of the *CIP Client* which provides functionally for local data management, e.g. saving a result set, or converting result sets into orders.
CIP Client Presentation Layer	Part of the *CIP Client* dealing with how information is presented to the human user, including all issues related to HMI, as well as dealing with certain format specific issues, e.g., displaying browse imagery.
CIP Operation	Based on Z39.50 definition, an initiating CIP request message and the corresponding terminating response, along with intervening related messages. Multiple operations may occur within a *CIP Session*.
CIP Session	A set of CIP messages exchanged between an *Origin* and *Target* beginning with an initialization message and ending with a close message between which the *Origin* and *Target* maintain state information concerning the interaction. (A *CIP Session* is a Z39.50 Z-Association.)
CIP Message	A unit of information transferred between an origin and a target whose format is specified as a Z39.50 Application Protocol Data Unit (APDU) possibly containing CIP specific APDUs in the external portion of a Z39.50 APDU.

Collection	A grouping of item descriptors that have commonality. A collection consists of a number of attributes that describe the collective contents of the collection, the values of these attributes can then be searched on to select items of interest to the user. Collections also have members; these are the unique identifiers of the items that are grouped by the collection rather than their collective descriptions. As collection members can be identifiers of other collections, a hierarchy of collections and product/guide members can be established, therefore permitting a flexible and powerful organization of data.
Collection Management Tool	Used by the RMA for tasks involved with populating and maintaining the data in the Retrieval Manager. These tasks involve translating collection or directory information into CIP collection format and checking for valid entries.
Existing Agency Client	Software elements which interact with a data providers *Catalogue* using the data providers protocol.
Guide data	Data that is available to the user to enhance understanding of the EO data, spacecraft, instrument, etc., and hence make a detailed analysis of whether the product data will be of value for a particular application. Guide data may also contain information necessary for processing the product data further.
ICS Client	Component of ICS that provides user access to CIP services via a CIP Client, and access to ICS Guide System services via an HTTP Client.
ICS Gateway	A software element which provides non-ICS clients access to ICS catalogues, e.g., searches by *Existing Agency Clients* to a *Catalogue* are sent to a *Retrieval Manager* using an *ICS Gateway.*
Item Descriptor	Used by CIP to represent items. The descriptor is a set of attributes. Item descriptors are of the following types: product, collection, guide, user.
Monitoring and Control Tools	Provides the machine-to-machine interface for integrating the operations of the *Retrieval Manager* with the operations of SSM.
OHS Translator	One of three types of ICS *Translators. OHS Translator* converts CIP messages into a data providers protocol for the purposes of order specification, quotation, and request.
Order Handling System	That part of a data provider which provides services associated with ordering products and guide.
Origin	Based on the Z39.50 definition for origin, where an *Origin* is that portion of a client or middleware which may initiate a *CIP session* with a *Target*. ICS elements capable of acting as *Origins* are *CIP Clients, Retrieval Managers, ICS Gateway* and *HTTP/CIP Gateways*.
Other CIP Based Federations	Groups of organizations other than CEOS using CIP to provide catalogue interoperability to data providers. These providers are interoperable with CEOS using CIP. These other federations may or may not use the ICS system design as a basis of the federation design.
Other Z39.50 Based Services.	Catalogue interoperability services provided using Z39.50 Version 3 but not necessarily CIP.
Primary Order	A CIP order between a *CIP Client* and a *Retrieval Manager*. (See *Secondary Order*)
Product data	A unique aggregation of data generated from information held in, or to be held in an archive (for predicted products). It can be located and retrieved by a user via CIP, possibly following further processing, such as map projection, sub-setting, band selection, etc., after or during extraction of the raw data as stored in the archive.

Registered Collections	Are owned by the EO data provider and described by a collection descriptor. These collections support the full range of CIP access services including discovery, navigation, location and searching. There are two types of registered collections which are distinguished by the purpose for which they were created, Archive Collections and Theme Collections.
Retrieval Manager	A *Retrieval Manager* services (and may be installed at) each catalogue site, it is used to integrate together the local catalogue systems and provide communication between users and other catalogue site *Retrieval Managers*. It is anticipated that each catalogue site is at least connected to one *Retrieval Manager* and that *Retrieval Manager* would 'know about' or 'own' a number of collections. The data within these collections would be the responsibility of that *Retrieval Manager*, with external collections referenced only and managed by their respective *Retrieval Managers*. The *Retrieval Managers* at each catalogue site would also communicate with each other using the CIP. The *Retrieval Manager* would then also communicate with local catalogue servers, such as archives and inventories, within its own site to services requests received from users. Another key function of the *Retrieval Manager* is to route search queries to other relevant *Retrieval Managers* and consolidate the search results before returning them to the user.
Retrieval Manager Administrator	The human operator that performs all tasks needed to establish and maintain a *Retrieval Manager*. In practice this is more than one person as the tasks are various types: scientist for collection definition, data base expert for maintaining CDB, system operator for diagnosing and correcting operational activities, etc. For convenience purposes all of these tasks are performed by the *RMA*.
Secondary Order	A CIP order is created by a *Retrieval Manager* in response to a *primary order* and may be either to another *Retrieval Manager* or to an *OHS Translator*
Site System Management	That part of a data provider which provides coordinated, on-line management of the distributed processing environment.
Target	Based on the Z39.50 definition for target, where a *Target* is that portion of middleware and servers which accept requests for *CIP sessions* from an *Origin*. ICS elements capable of acting as *Targets* are *Retrieval Managers* and *Translators*.
Task Package	The set of attributes that describe an activity which is started by an Extended Services Request. Based on Z39.50 definition for a Task Package.
Theme Collections	This type of collection may be set up by data providers or users who want to organize some of their data into groupings which differ from their provider archive collections (i.e. from the baseline inventory), for the convenience of their users, for example, based on the geographical area covered, the scientific discipline supported by the data, the instrument type. Etc.
Translators	Software element which converts CIP into the protocols used by a data provider. Three *Translators* are identified in ICS: *Catalogue Translator, OHS Translator, UPS Translator.*
UPS Translator	One of three types of ICS *Translators. UPS Translator* converts CIP messages into a data providers protocol for the user information, e.g., address, and for authentication.
Unregistered Collections	Are likely to be created and owned by an end user of EO products that has created and populated the collection to obtain a single source of thematic information. This will then enable further analysis or easy access by themselves or other users. These are collections of potentially quite disparate item descriptors of interest to a relatively small user community researching a particular theme, i.e. in the example, the mid-west flood of 1993. Unregistered collections need not be described by a collection descriptor and cannot be discovered in a search of ICS collections.

User Profile System	That part of a data provider which provides services associated with user information and authentication.
User Session	A user session represents the interaction between a human user and the CIP (i.e. a Retrieval Manager), which has in general been established by the user via a terminal which may be running a WWW client or a specific CIP client Man-Machine Interface. This, within CIP-B, will be an authenticated session, e.g. with the transmission and acceptance of a user name and password. The term 'log on' is used within this document to mean the establishing of user interaction session with a catalogue system supported within the CIP domain. In this context, the word 'session' should not be interpreted in terms of communications sessions or states. A user interaction session is independent of the underlying communications layer. Note further, that as physical communications layers can be broken, it is likely that a user identifier will need to be retained within the CIP domain and therefore, user identifiers will be exchanged.
Users	The user represents the combination of a real human user and the client software that the human user is using to interface with the Retrieval Manager. The CIP is not concerned directly with the client software, although the CIP has to be able to support the tasks that the user wishes to achieve, and it is anticipated that in general, the actual CIP should be transparent to the user. This is analogous to a human not directly interacting with TCP/IP, but being aware that it satisfies the task of data transfer. The only exceptions to this may be when the user is controlling a query or when error or state information is generated by the CIP under anomalous circumstances. The user has a number of important properties, such as a unique identifier, option, etc. (of course this does not exclude an anonymous user having a set of default properties). Generally, the user can perform three types of tasks, either place a query or place an order or control a session. Note that there is also a special type of user in the CIP domain and that is a non≠ human user such as another Retrieval Manager or a scheduler. These could for example place orders as required, without human interaction, apart from the initial set-up of the schedule.

1.6 References

ICS Compatability: Explanatory

[R1]∞ *Protocol Task Team (PTT) Development Plan*, CEOS/WGISS/PTT/Plan, Issue 1.0, 19 July 1996, Committee on Earth Observation Satellites /Protocol Task Team

[R2] ∞ *Interoperable Catalogue System (ICS) User Requirements Document (URD)*, CEOS/WGISS/PTT/ICS-URD, Issue 2.2, March 1997, Committee on Earth Observation Satellites /Protocol Task Team

[R3] ∞ *Catalogue Interoperability Protocol (CIP) Specification - Release B*, CEOS/WGISS/PTT/CIP, Issue 2.3, to be ratified February 1998, Committee on Earth Observation Satellites /Protocol Task Team

[R4]∞ *ICS Valids Document*, April 1998, Committee on Earth Observation Satellites /Protocol Task Team

[R5]∞ *ICS Collection Manual*, CEOS/WGISS/PTT/CM, Issue 1.0, April 1997, Committee on Earth Observation Satellites /Protocol Task Team

[R6]∞ *Technical Architecture Framework for Information Management (TAFIM), Volume 3: Architecture Concepts and Design Guidance*, Version 2.0, 30 June 1994, US Defense Information Systems Agency Center for Architecture.

[R7]∞ *Summary of the ECS System Design Specification*, ECS Technical Paper Number 19400436, June 1994, Hughes Information Technology Systems (HITS) ECS Project Office. (document is available at http://edhs1.gsfc.nasa.gov/)

[R8]∞ *CEOSnet Resource Planning and Coordination Document*, CEOS Network Subgroup, Draft Version 0.4, April 1997

[R9]∞ *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*, ANSI/NISO Z39.50-1995, Official Text, July 1995, Z39.50 Maintenance Agency.

[R10]∞*DPRS Technical Note/002*, LUK.502.EC21317/TN002, Issue 0.1, 20 September 1995, Logica UK Ltd.

[R11]∞*ICS Collection Technical Note*, CEOS/WGISS/PTT/CTN, Version 1.0, 8 July 1996, Committee on Earth Observation Satellites, Protocol Task Team

[R12]∞*Internetworking with TCP/IP: Volume 1, Principles, Protocols and Architecture*, Douglas E. Comer, Prentice-Hall, 1995.

[R13]∞*Z39.50 Information Retrieval Protocol in the Internet Environment*, Clifford A. Lynch . (Document is available at http://lcweb.loc.gov/z3950/agency/)

[R14]∞*ISO 8824 - Information Processing Systems* - Open Systems Interconnection - Specifications for Abstract Syntax Notation One (ASN.1), 1987

[R15]∞*ISO 8825 Information Processing Systems* - Open Systems Interconnection- Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1) 1987

[R16] *RFC 1340 Assigned Numbers*. Reynolds, J.K.; Postel, J.B. July 1992

[R17]∞*CEO Enabling Services Usage Scenarios Document*, CEO/ES/WP 2200/SD-1/184, Document Issue 1, July 1996, Centre for Earth Observation (CEO).

[R18]∞*Hypertext Transfer Protocol -- HTTP/1.0 Specification*, RFC Number 1945, May 1996, T. Berners-Lee, R. Fielding, and H. Frysty.

[R19]∞*ICS Query Performance Estimation, NASA ECS White Paper*, Document 170-WP-014-001 Hughes Information Technology Systems, August 1997.

[R20]∞*Internet Official Protocol Standards,* Internet Architecture Board Standards Track, Document RFC #2200, Standard #1, J. Postel, Editor, June 1997

[R21] *Queuing Systems*, Volume 1: Theory, Leonard Kleinrock, John Wiley, 1975.

[R22]∞*ECS User Scenario Functional Analysis*, NASA ECS White Paper, Document 194-00548TPW, Hughes Information Technology Systems, October 1994

[R23]∞*The Directory - ITU-T Authentication Framework*, ITU-T (formerly CCITT), Recommendation X.509, 1994 (Version 3 extensions)

[R24] ∞*Interoperable Catalogue System (ICS) Guide Design and Protocol Specification*, CEOS/WGISS/PTT/GDPS, Issue 1.0, January 20, 1998, Committee on Earth Observation Satellites /Protocol Task Team

[R25]∞*Z39.50 Application Profile for the Content Specification for Digital Geospatial Metadata or "GEO"*, Verssion 2.1, April 1998, Federal Geographic Data Committee

[R26]∞*Transmission Control Protocol*, IETF Document RFC #793, Standard #7, DARPA Internet Program, September 1981

[R27]∞*Network Security CEOS Context*, CEOS Network Subgroup, (Annex 1 by Grant Miller) September 1997.

[R28]∞*Order Options Amendment to the CIP Specification*, CEOS/WGISS/PTT/CIP-ORD, Planned to be released October 1998, Committee in Earth Observation Satellites

This page intentionally left blank.

2. CATALOGUE INTEROPERABILITY

This section provides an overview of key issues in catalogue interoperability. The following topics are addressed in this section:

- Purpose and scope of catalogue interoperability
- ICS Concepts
  - Collections data model
  - CIP as a Z39.50 profile
  - Browse data in CIP
  - Product ordering and security
  - Guide Documents in ICS
- Levels of Compliance to CIP, ICS and IGP

2.1 Purpose and Scope of Catalogue Interoperability

ICS Compatibility: Explanatory

The Committee on Earth Observation Satellites (CEOS) is comprised of international space agencies. CEOS promotes the interoperability of space agency catalogues through the definition and development of interoperability concepts. By enhancing the standardization of EO data and information management services, CEOS enables the catalogue services to be more accessible and usable to data providers and data users world wide. EO catalogues services, as defined by CEOS, are as follows:

- search and retrieval of information about EO data products
- order of EO data products
- searches and retrieval of Guide documents that complement the EO data products

Catalogue Interoperability in this context is defined as: the ability to provide a Data User with the appearance of a single, unified catalogue for all participating data providers. In order to provide catalogue interoperability all participating data suppliers must support at least one common method(i.e., API) for accessing functions such as authentication, directory, inventory, guide and order. Each supplier may support additional consumer functional interfaces to support their private data users

Catalogue interoperability may extend beyond just the members of CEOS in promoting data access within a wider community of EO data providers and eventually to non EO data providers.

2.2 ICS Concepts

This section introduces key concepts for the understanding of the CIP, IGP and ICS.

2.2.1 Design Approach: CIP Space, IGP Space and ICS

ICS Compatibility: Explanatory

The PTT design approach considers catalogue interoperability as the loose coupling of a federation of existing catalogue systems using a set of common protocols. The approach provides users the services available at all sites regardless of which site the user established a connection with.

- The Catalogue Interoperability Protocol (CIP) standardizes the services needed for interaction between users and catalogues of EO data products
- The ICS Guide Protocol (IGP) standardizes the services needed for a user to discover EO Related documents (i.e., guide documents).
- The Interoperable Catalogue System (ICS) is a design that uses CIP and IGP as the common protocols between data providers and users of the data.

The objective of implementing CIP and ICS is to provide more users with access to more data more easily. The ICS domain can be seen in Figure 2-1 as divided into two virtual domains;

- 'CIP domain' within which CIP messages, consisting of requests and responses, are exchanged between architectural elements.
- 'IGP domain' within which IGP messages, consisting of requests and responses, are exchanged between architectural elements.

These two domains are separable and can exist independently from each other. To allow access to both domains, an ICS client was designed with a CIP Client component and an IGP Client Component. To enable consistency in the ICS domain a Collections Management Tool (CMT) was defined to update the contents of both the CIP and IGP domains simultaneously.

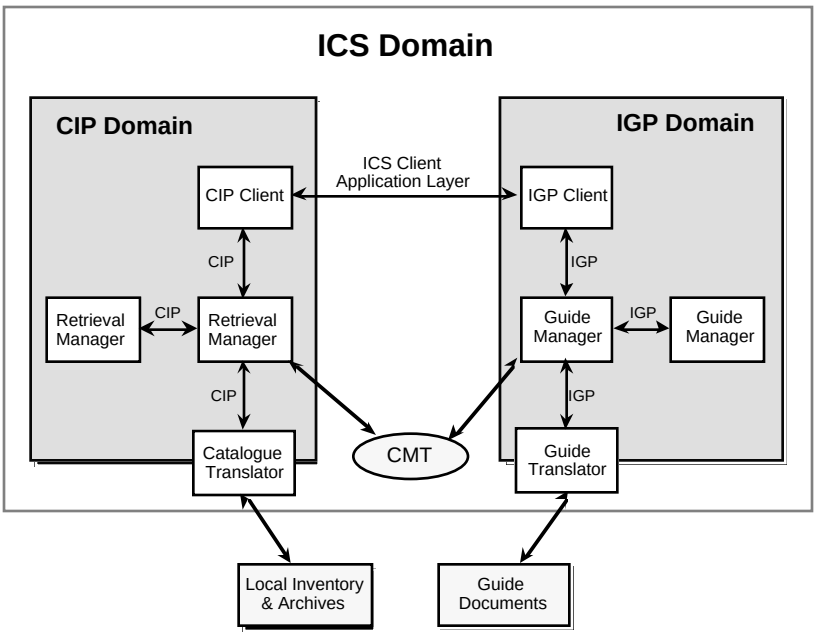

Figure 2-1. ICS Domain

To support transparent access to multiple catalogues, a three tier structure was used to design the ICS space. Clients exchange messages with a middleware layer which in turn interacts with multiple catalogue servers. The middleware provides the routing and translation services to allow client requests to be presented at the multiple heterogeneous catalogues. The middleware is of two types of elements: Managers and Translators. Managers provide an access point for clients and route the requests to the various servers. Translators, bound with the clients and servers, translate CIP or IGP to and from the native protocol of the client or server. Future client and server developments may use CIP or IGP directly and hence not require translators

This approach supports a diversity of clients, and servers. Clients may be used directly by a human user or may be an agency system acting on behalf of a user. Depending on the design of an existing catalogue system, services may be provided by different servers and translators. Because the routing service provided by the Middleware is independent of the type of service, separate translators may be provided for inventory, browse, ordering, and user profiles. This architecture is also applicable for small data providers, such as university research groups, who are unable to provide adequate middleware at their site but still wish to join the ICS domain. Their local catalogue, inventory and guide documents can be made available to the ICS community by the inclusion of appropriate descriptions within another agency's middleware.

CIP Space is a protocol centric view of catalogue interoperability and provides for the loosest coupling needed to achieve catalogue interoperability among a wide community of EO data providers and users of EO data. A range of design solutions is permitted by the CIP and IGP spaces. To provide for a higher degree of uniform services at the cost of additional agreements between agencies, additional design criteria for interoperability are defined in the ICS design document. The additional design definitions pertain to the allocation of functionality and data amongst components, agreement on an underlying communication protocol, and agreement on how to conduct distributed system management of ICS. The difference between CIP Space and the ICS is depicted in Figure 2-2. CIP Space is defined by those CEOS agencies and other federations and organizations which provide catalogue services using CIP and/or guide service using IGP. Those CEOS agencies which provide services, communications and systems management compatible with the ICS design make up the ICS. Its should be noted that while all ICS members must implement CIP, guide handling is considered an optional element of ICS and an ICS members may choose not to implement IGP. Note that other federations may choose to use the ICS design as the basis for their federation.

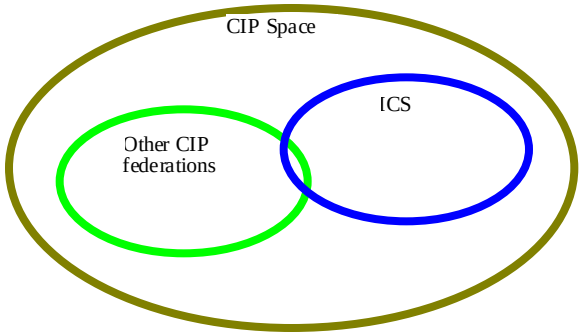

Figure 2-2. VENN diagram of CIP Space and ICS

Assuming query and result routing between geographically dispersed sites (see Figure 2-1), an agreed middleware layer and its interfaces to users and providers needs to be in place. To define such a system, the PTT have established the following CIP, IGP and ICS standards:

- Search Standardization - the functions and procedures how search can be invoked and executed both by end user clients and middleware; the search language and it's syntactical rules;
- Retrieval Standardization - the procedure for retrieval of query results and other objects which need to be presented to the client; an exact description of what is the retrieval content, its format and meaning.
- Attribute Standardization - the list of fields or attributes which can be searched and the definition of their semantic meaning. In addition CIP provides a local attribute mechanism which allows the data producer to extend the set of attributes to include more information to assist the user in the selection of EO products of interest.

- ↑ Dynamic Client Configuration Standardization - a mechanism which is used for dynamic client configuration, i.e. to make a client understand which functionality can be invoked at the server side, which attributes are understood, and what are their meanings.
- ↑ Order Procedures Standardization - a common way of defining product lists and associated order, packaging and online/offline delivery options; communication of price and accounting information; order and delivery addresses.
- ↑ Security Policies Standardization - a reliable and secure mechanism for user authentication and authorization.
- ↑ Guide Document Discovery and Retrieval Standardization - a mechanism which allows full text or fielded searches on documents relating to EO data holdings. This mechanism uses HTTP protocols and technology to enable the use of standard web browsers and discovery services such as Alta Vista to discover and retrieve EO related documents of interest. This service provides the general WWW community access to documents describing EO data. This service also provides pointers to EO data related to each document.

2.2.2 Collections Data Model

ICS Compatibility: Explanatory

In an interoperable catalogue environment it is essential to organize metadata by distinguishing user and provider views as well as archive-oriented and theme-oriented structures. It is important to define a mapping between the different views and structures which often will lead to a hierarchical relationship of collections. The ICS data model is based on the notion of collections. A collection may contain descriptors for data products or descriptors for other collections. In addition to the value of collections for presentation of data organization to users, collections provide the mechanism for routing distributed searches. When a collection contains both local and remote members, the Retrieval Manager may search the local site as well as sending the search on to the remote site.

The collection concept is visualized in Figure 2-3 below. The collections in the diagram are numbered so that their relationship can be easily seen; they do not represent the naming of collections in an actual implementation. The terminal collections (labeled '1.x') group the product descriptors as is appropriate. As can be seen the collections can overlap each other and product descriptors can appear in more than one collection. Above the terminal level collections, there are non-terminal collections that group together any number of other collections. The grouped collections do not all have to be at the same hierarchical level and this grouping of collections can continue to any hierarchical level, with existing collections being included at any other arbitrary level. A non-terminal collection could group together terminal collections and other non-terminal collections (as the link between collections 3.1 and 1.5 shows). Also, a terminal collection could exist without a relationship to a higher collection (i.e. collection 1.9), or a non-terminal collection

could exist with no relationship to lower collections, in other words a collection without members (i.e. collection 2.5). Collection 1.9 can not be reached by a hierarchical search, but could be located if its URL was made public (an example of such a collection may be a persistent result set or a collection under construction).

Collections can be used to group data together which have a similar semantic theme. All collections support the search mechanisms defined in the CIP. CIP defines two types of searches which a CIP user may request:

- Collection Search: finds collections of interest without searching collections containing products
- Product Search: finds individual product descriptors which may eventually lead to the order of an actual product.

Additionally, the user may request that the search be contained locally to the target Retrieval Manager (i.e., a local search), or request that the search be propagated to other Retrieval Managers based on the collections (i.e., a distributed search).

Two types of collections are defined: registered collections and unregistered collections.

- *Registered collections* are owned by the EO data provider and described by a collection descriptor. These collections support the full range of CIP access services including discovery, navigation, location and searching. There are two types of registered collections which are distinguished by the purpose for which the were created: Archive collections and Theme collections.
  - Archive collections: : This type of collection is likely to be created by data providers to organize their archives and facilitate access to the product descriptors (i.e. analogous to an inventory containing inventory entries
  - Theme collections: This type of collection may be set up by data providers or users who want to organize some of their data into groupings which differ from their provider archive collections (i.e. from the baseline inventory), for the convenience of their users, for example, based on the geographical area covered, the scientific discipline supported by the data, the instrument type, etc.

- *Unregistered collections* are likely to be created and owned by an end user of EO products that has created and populated the collection to obtain a single source of thematic information. This will then enable further analysis or easy access by themselves or other users. These are collections of potentially quite disparate item descriptors of interest to a relatively small user community researching a particular theme, i.e. in the example, the mid≠ west flood of 1993. Unregistered collections need not be described by a collection descriptor and their collection attributes cannot be discovered in a search of ICS collections.

An unregistered collection may be registered by an ICS data provider. The process of registration may vary significantly between ICS data providers, but minimally requires the creation of a collection descriptor and the insertion of that collection descriptor into the ICS data providers collection structure. Generally the registration of a collection will also involve a scientific review and the transfer of ownership of the collection from the end user to the ICS data provider to ensure accuracy and long-term availability of the registered collection. Further description of the registration process can be found in the ICS Collections Manual [R5].

Note that these category definitions are not mandatory for the CIP to operate, but help to distinguish collection categories for discussion purposes. The CIP does not distinguish between the categories (the Retrieval Manager does, however) and the same CIP search and retrieval services are applicable to all collections. The Retrieval Manager does make use of collections for routing of distributed queries. Standardization of collection definitions is provided as part of the ICS design.

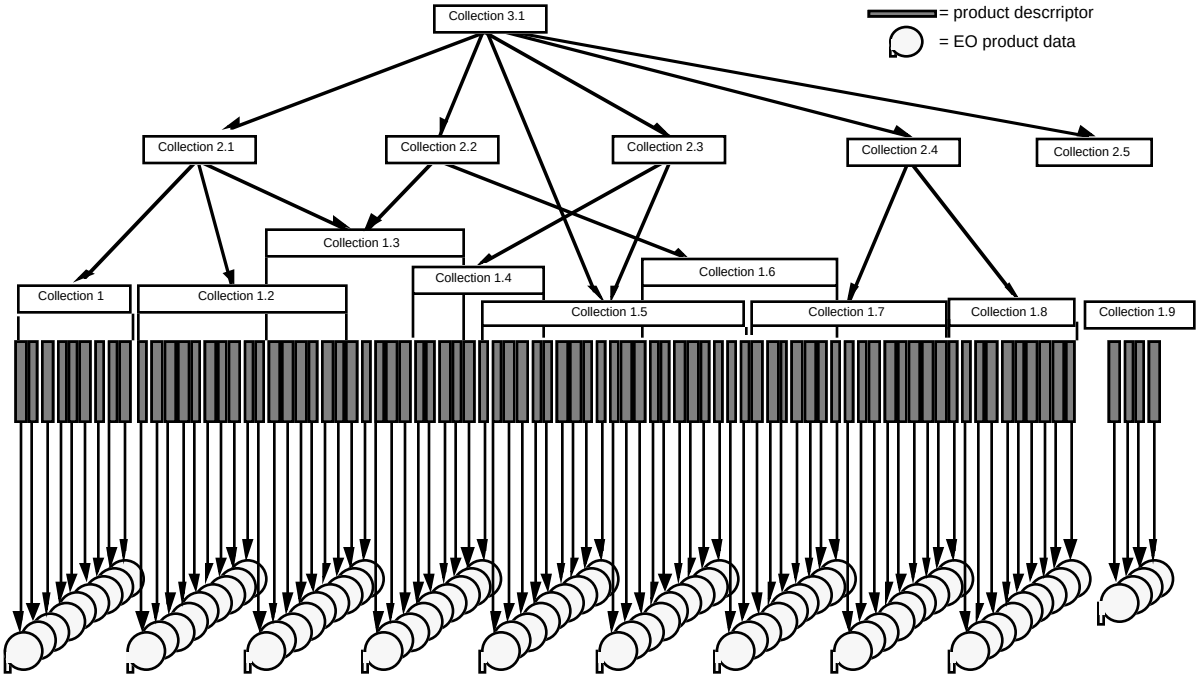

Figure 2-3. The Concept Of a Collections

2.2.3 CIP as a Z39.50 Profile

ICS Compatibility: Explanatory

Based on a set of user requirements and an analysis of existing communication standards, Z39.50 was selected as the base protocol for CIP. The Z39.50 protocol [R9] is designed for information search and retrieval within a generic domain which, together with the powerful services and data structures it supports, makes it an ideal basis of an EO domain search and retrieval protocol.

CIP has exploited and extended the services of Z39.50 to provide distributed searching, extensions to attribute set definitions, and the definition of a secure ordering service. CIP is a profile of Z39.50, i.e. it defines the use of the Z39.50 facilities within the CIP domain and defines the attributes that are used to search and present EO information. Other Z39.50 profiles include GILS, GEO and the Digital Collections Profile. CIP extends Z39.50 for distributed searching by supporting the collection data model discussed in Section 2.2.2 which allows hierarchies of related collections to be constructed and searched.

The GEO Profile supports Geographic Information Systems (GIS) applications and thus is of special interest to users of EO data. For this reason an alignment of the CIP and GEO profiles was made. The objective of this alignment was to allow both GEO and CIP clients to search and retrieve records from databases defined by either profile, and thereby maximize interoperability. The alignment was helped by the similarity of the spatial and temporal attributes of the metadata, but needed to take into account the different data models in CIP and GEO. It should be emphasized that the CIP/GEO interoperability is for search on the intersection of CIP and GEO attributes and the retrieval of item descriptors. There is no interoperability on the more advanced functions of CIP such as ordering and security. Additional support for compatibility is provided by the requirement that Retrieval Managers must support access by any Z39.50 Version 2/3 compatible client.

2.2.4 Browse Data in CIP

ICS Compatibility: Explanatory

Browse data helps users to evaluate EO products. Browse data are typically reduced resolution or summary data versions derived from the EO product data itself. Browse data are delivered to the user via two different mechanisms, dynamically over the network during a user query session, and as an EO product order. The second case allows users to order the Browse data from an archive system to be delivered separately from their query session. This means that the user can then store and access the data locally rather than dynamically over a network. It is important to

note that although most catalogue systems will provide some form of reduced data retrieval, it is not a mandatory CIP service. The form and content of browse data is dependent on the nature of the associated EO data and the data selection criteria necessary for a science discipline to evaluate the EO data. Browse data in the CIP is seen as one of the following forms:

- Browse attribute - simple attribute containing the actual Browse data.
- Browse compound - compound attribute containing attributes describing the Browse data, including the simple Browse data object.

2.2.5 Product Ordering and Security

ICS Compatibility: Explanatory

CIP supports a wide range of ordering services, e.g. the specification of order options and provisions for authentication and non-repudiation of orders. A user can retrieve the order options associated with a product, where order options may be processing as well as packaging options. CIP allows a local order handling system to define or refer to its own local order options. But the CIP also contains a standard mechanism to define order options..

A user can request a quote for a specific order and submit the order. The order process is monitored by the Retrieval Manager and can be queried later by the user to determine the status of the order. To support ordering of data for which a user must have privileges or for orders which the user will be charged, a authentication scheme has been defined. The authentication supports digital signatures using either a shared (symmetric) key approach or an public (asymmetric) key approach. Authentication allows the Retrieval Manager to identify the user with an appropriate level of confidence and enables the Retrieval Manager to log the authenticated user requests to provide non-repudiation. The CIP security approach avoids the need to transfer password information over the network. Future enhancements to CIP anticipate the ability to support the transfer of financial information to support billing.

If a provider decides to create a collection containing browse products, a user would be able to order browse products via the CIP. In this case the browse product would be considered as any other data product and not any longer as “browse” as defined in the previous section.

2.2.6 Guide Documents in ICS

ICS Compatibility: Explanatory

Much of the metadata for EO data collections is not easily stored in a structured form. This information is stored in documents called Guides. Since Guide documents provide information that is required for the understanding of some EO data collections, they must easily be accessed via ICS mechanisms that provide search and retrieval of catalogs. Guide documents also provide a human readable descriptions of EO data collections and are often used by new EO data users as a discovery mechanism to identify collections of interest. It is the goal of the ICS to make this

discovery mechanism as simple and widely available as possible to extend the uses of EO data to communities which have not traditionally used EO data. This goal has resulted in the definition of a guide system in ICS which uses ICS Guide Protocols (IGP) based on HTTP and enables general purpose Internet Search/Discovery Engines such as Alta Vista to locate EO Guide documents either by free text or attribute value searches.

This system is not based on Z39.50 and is not a mandatory capability of an ICS node. However there is a strong linkage between the CIP client/retrieval manager and the HTTP based client and indexing method for Guide Documents. To allow coordinated access to catalogues and documents an ICS client was designed with a CIP Client component and an IGP Client Component. The ingest of documents and collections into the ICS is coordinated by the Collection Management Tool (CMT) to assist in maintaining the consistency of the collection descriptor and the HTTP index that enable search and access of Guide Documents. Further details of this ingest process are discussed in the ICS Collections Manual [R5]. The specific design of the guide system can be found in the ICS Guide Design and Protocol Specification [R24].

2.3 Levels of Compliance to CIP, IGP and ICS

ICS Compatibility: Explanatory

The ICS SDD and CIP Specification are detailed documents with many services and mechanisms specified. As discussed in Section 2.1, agencies may choose to implement a wide range of these services in their CIP Clients and Retrieval Managers. It is critical for the designers and implementors of these software components to understand what capabilities are critical to the minimal operations of the ICS and must be implemented in all components versus those capabilities which are optional. In addition it is assumed that various CIP based components will be available either as shareware or commercial software. The developers of ICS or other CIP federations will need a method to categorize and select among these available components. For this reason, compliance levels have been defined within both the ICS SDD and the CIP Specification.

- *ICS Compliance*, which is defined in Section 9 of the ICS SDD, discusses the required minimum configuration for a site which wishes to be considered ICS compatible. The minimum site is based on the CEOS policy.
- *CIP Compliance*, which is defined in the CIP Specification, discusses specific CIP messages and parameters must be supported by a CIP Client or an RM.

These compliance concepts are interdependent since in order to support a specific ICS service, the RM and CIP Client must support the CIP messages which enable that service.

IGP has only one compliance level which is full compliance so no system that does not implement the full IGP Specification can be considered IGP compliant.